ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-MOPHA112

IMPROVING PERFORMANCE OF THE MTCA SYSTEM BY USE OF PCI EXPRESS NON-TRANSPARENT BRIDGING AND POINT-TO-POINT PCI EXPRESS TRANSACTIONS

L.Petrosyan[*], DESY, Hamburg, Germany

Abstract

Increasing Performance of the MTCA System by use of PCI Express Non-Transparent Bridging and Point-To-Point PCI Express Transactions.

The PCI Express provides one of the highest data transfer rates today. However, with increase in number of modules in a MTCA crate and client programs, and also with complication of modules and increase in number of module registers, as well with increase in amount of data requested by the users the performance of the whole system plays an important role.

Distributed systems are gaining popularity as they fill the need of next generation systems.

Multiprocessor systems provide not only the ability to increase processing bandwidth, but also allow greater system reliability through host failover.

The use of non-transparent bridges in PCI systems to support intelligent adapters in enterprise systems and multiple processors in embedded systems is well established. In these systems, the non-transparent bridge functions as gateway between the local subsystem and the backplane. Such applications can be ported to PCI Express by the use of non-transparent bridges, with the non-transparent bridge integrated into a PCI Express switch in place of one of the transparent bridges.

PCIE NON-TRANSPARENT BRIDGE

MTCA systems use the PCIe as a central bus. Adding a second CPU to the existing MTCA system, using Non-Transparent Bridging, will allow to increase the performance of the system.

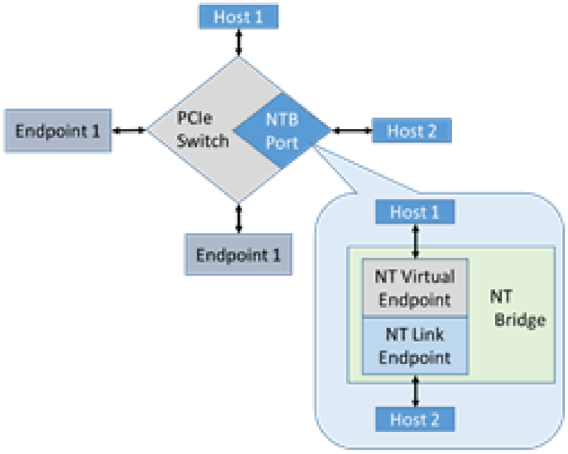

Figure 1: Non-Transparent Bridge.

PCIe has to have only one Root Complex, PCIe Bus configuration and memory mappings has to be done by the one Host to avoid the Bus numbering and memory mixing.

Non-Transparent Bridging used to connect two independent address/Host domains; it allows to connect second Host to the existing PCIe bus. A Non-Transparent Bridge consist of two back-to-back PCIe endpoints, a Virtual and Link side endpoints (Fig. 1). A Non-Transparent Bridge isolates the address spaces of different Hosts by appearing as an endpoint to each side.

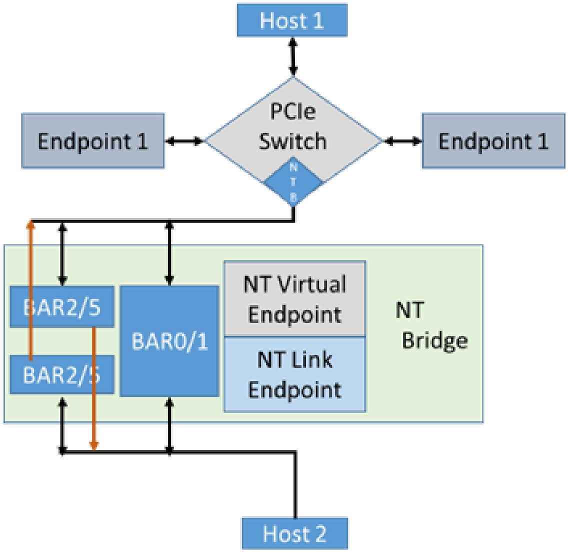

Figure 2: Key elements of the Non-Transparent Bridge.

The Key elements of the NTB (Fig. 2):

- BAR0/1 used for NTB configuration, visible from both sides of the NTB
- Up to 4 BARs, individually enabled
  - BAR 2/5 are aperture into the address space on the far side of the other endpoint, provides ad-dress transaction from one side to other
- 8 Scratchpad Registers (BAR0)
  - Provide a means of communication between two Hosts over a non-transparent bridge. They are readable and writeable from both sides of the NTB
- 16 Doorbell Registers (BAR0)
  - The doorbell registers are used to send inter-rupts from one side of the NTB to other.

The PCIe packets are routed by address and Bus number, so to send packets from one side to other the addresses and bus number have to be translated.

[*] ludwig.petrosyan@desy.de

The Packets passing through the non-transparent bridge provided by:

1. Direct Address Translation Register for each enabled BAR on both sides (Fig. 3.)
   - The content of the register could be the PCIe address of some endpoint in other side of the NTB.

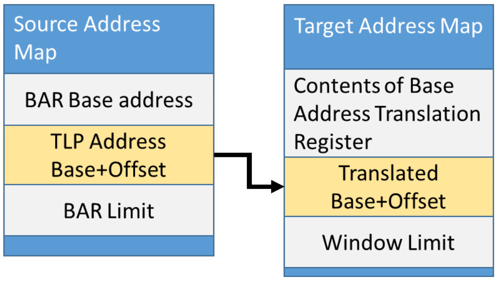

Figure 3: Direct Address Translation Register.

2. Requester ID conversion across NTB, Re-quester ID translation lookup tables (Fig. 4).

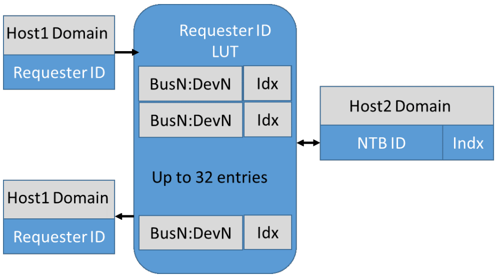

Figure 4: Requester ID translation lookup tables.

To enable and use of the NTB some configuration has to be done.

In case of MTCA system we have two possibilities to set the NTB:

1. Set the NTB port on the second CPU and insert this CPU to any slot of the MTCA crate,
2. Set one of the MTCA crate slots as NTB port and insert CPU in this Slot.

The second approach is better, we are independent from second CPU. In our test we used second approach (Fig. 5).

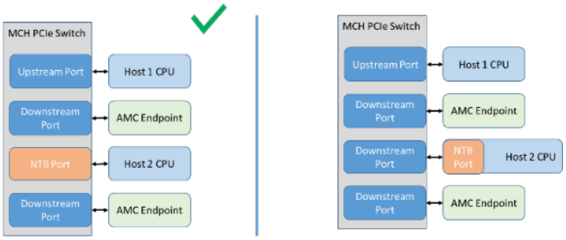

Figure 5: Two possibilities to set the NTB.

To enable and use the NTB some configuration has to be done. The configuration of the NTB could be dividev in two parts:

1. Boot Time configuration. Done by MCH. After the boot the NTB ready for use.
   a. Enable NTB on the Current port,
   b. Enable BARs and set Sizes for each BAR.
2. Run time configuration. Done for each transaction by NTB device driver.
   a. Setup Address Translation Register for current BAR,
   b. Setup Requester ID Look Up Tables,
   c. NTB Virtual and Link side device drivers communicate using Scratchpad Register and Doorbell Register to share the information. The drivers setup Address Translation Tables and Requester ID Look Up Tables. The drivers initiate PCIe Transactions.

PCIE NON-TRANSPARENT TEST

Boot Time-MCH: PCIe Switch on MCH configured to have 2 BARS on each NTB side with 1 MB size. Run Time, NTB Device Drivers on both sides.

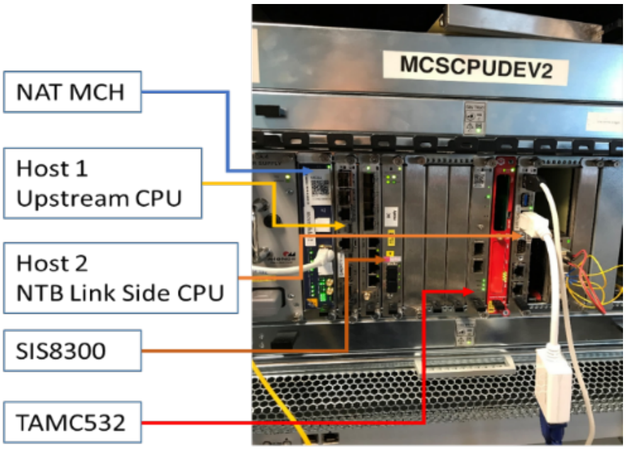

Figure 6: Transaction from Host2 to SIS8300 and TAMC 532 are tested.

NTB Virtual and Link side device drivers communicate using Scratchpad Register and Doorbell Register to share the information.

- Used Virtual and Link side NTB device drivers created on base of Universal driver
- Universal driver has information about all AMC PCIe endpoints
- Virtual/Link side Driver sets Address Translation Registers:
- Link Side BAR2 address of SIS8300
- Link Side BAR3 address of TAMC532
- Virtual Side Driver sets RID-LUT:
- Virtual Side LUT: Root Complex
- Virtual Side LUT: SIS8300
- Virtual Side LUT: TAMC532
- Link side Driver sets RID-LUT
- Link Side LUT: Root Complex

Transaction from Host2 to SIS8300 and TAMC532 as well communications throw the Scratchpad registers and interrupts throw the Doorbell registers are tested (Fig. 6).

USE CASES

Different Use Cases (Figs. 7 and 8):

- Send Data from one Host to other
- Link Side server controls one of AMC Module
- Virtual Side server controls both modules
- One module send Data to Virtual Side
- Other module sends Data to Link Side using PCIe Poin2 Point connections
- Connect two MTCA system, or external CPU to MTCA

CONCLUSION

This short document reports our experience of establishing non-transparent bridging in MTCA system.

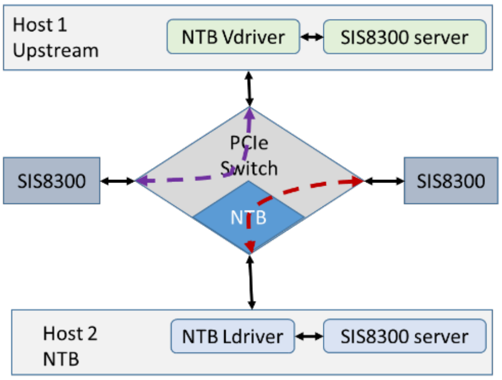

Figure 7: Link Side server controls one of AMC Module.

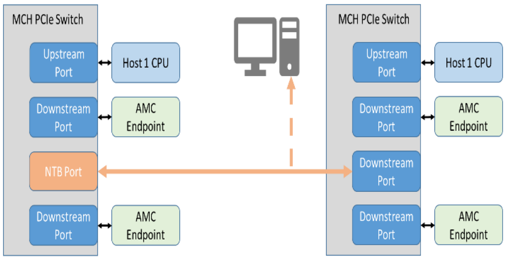

Figure 8: Virtual Side server controls both modules.